Effective Fuzzy Keyword Search Technique in Cloud

Atharav Arora
Student
Amity School of Engineering and Technology
Sector 125, Noida

Rushil Gupta
Student
Amity School of Engineering and Technology
Sector 125, Noida

Darothi Sarkar
Assistant Professor
Amity School of Engineering and Technology
Sector 125, Noida

ABSTRACT
This paper explains how an effective, still privacy preserving fuzzy keyword search in the domain of cloud based computing can be done.

The paper discusses the problem of extracting the file from a cloud server using minimal amount of resources. Whenever a user wants to find a file on the cloud server, he/she has to browse through a lot of files. It results in complete wastage of bandwidth and most important time. One of the easier and efficient way to do this is to retrieve the files through keyword-based search technique. By this, they won't have to retrieve files that are encrypted and doing such a thing would be totally impractical in case of cloud computing. So, the keyword based search allows the users to retrieve the files of their interest. Such a technique is applied widely in plain text searching scenarios, like search engines. But, data encryption prove to be a restriction to user's ability as it does not allow user to perform keyword search, making a plain-text-search method useless for cloud computing. [1]

An effective way is formalized for resolving the problem of performing an effective-keyword-search on the encrypted data stored in a cloud and yet, maintain the privacy of the keywords.

General Terms
Fuzzy keyword search, encryption, decryption, search, wildcard-based technique, web services.

Keywords
Fuzzy keyword search, Cloud based computing, encrypted data, privacy of keyword.

1. INTRODUCTION
Due to advancement in technology, an outburst of data is seen. To meet this never ending need of storage space, cloud computing was invented. A cloud storage could be thought of as a type of storage in which data is stored, maintained, managed and backed up over a network and is available to the users when and where they need it. With time, the prevalence of cloud computing has also grown. Now, most of the data including sensitive information like financial details, government documents, confidential material, etc. after storing all the data on the cloud, owners become relieved from the tension regarding maintenance and management of the data and enjoy an easy on demand access to their data anywhere, anytime. But, the fact is that they are unaware of the risk that the data they updated is not safe as he server could not be completely trusted after constant attacks from intruders. Hence, the information must be encrypted prior to outsourcing it so that the data remains confidential and its integrity is maintained.

Some users might want to outsource data to others. Hence, users might only want to retrieve only some files that they are interested in, rather than having a look at all the files. One of the easier and efficient way to do this is to retrieve the files through keyword-based search technique. By this, they won't have to retrieve files that are encrypted and doing such a thing would be totally impractical in case of cloud computing. So, the keyword based search allows the users to retrieve the files of their interest. Such a technique is applied widely in plain text searching scenarios, like search engines. But, data encryption prove to be a restriction to user's ability as it does not allow user to perform keyword search, making a plain-text-search method useless for cloud computing. Encryption also needs the keyword to be kept private because the keywords also contain some useful information about the data files. Hence, such an encryption can preserve the privacy and also removes any possible use of plain-text-search techniques.

2. BACKGROUND INFORMATION
The focus of this project is on enabling an effective, still privacy preserving fuzzy-keyword-search in the domain of cloud based computing.

In current scenario, a plain text search technique is used to find files over a cloud server. Such a technique is not only inefficient but also is unable to search over encrypted data on the cloud. Hence use of this technique is completely impractical in the case of cloud computing [2]. So to tackle this situation, we have formalized an effective way of resolving the problem of performing an effective-keyword-search on the encrypted data stored in a cloud and yet, maintain the privacy of the keywords.

Such a technique enhances the usability of the system by returning the files matching the keywords entered by the user to the predefined keywords or to the nearest possible files based on the keyword similarity semantics, only when there is no exact match found. The technique we would be using is to edit the distance so that we can quantify the keywords and develop a technique, like the wildcard based technique for generating the fuzzy keyword sets.

Also, it eliminates the requirement of enumeration of all the fuzzy keyword sets resulting in a minimal size of the keyword sets. In this way, in view of the developed fuzzy-keyword sets, we have proposed a powerful, yet proficient method for performing catchphrase seek. Furthermore through particular examination, we would be demonstrating that the proposed arrangement is totally secure and jam the security of the

information content but does not baffle in finishing its objective of a proficient fuzzy-keyword seek.

3. DESIGN METHODOLOGY

3.1 System Model

In this paper, a cloud data system is taken into consideration which constitutes a cloud owner, data users and a cloud server on which the files are stored. A user can upload as many files he/she wants and then the files are saved by a specific keyword referred as the 'fuzzy keyword' and then can search the files by the help of that particular keyword. Rule formulated for the above statement is that given that there is a collection of k encrypted files as D= {F1, F2, F3…… FK} in the cloud server, then there will be a set of distinct keywords for each file that could be formulated as W= {W1, W2, W3……. WQ}. The cloud server only allows the authentic and authorized. The user is authorized by the help of a username and password which he/she sets while registering for using the cloud server. The user then can either upload the files or view previously uploaded files. For uploading the files, the user has to first login the uploading account and then upload the files. For searching a pre-existing file, the user must be aware of the keyword of the file. He/she has to search the file with the help of that keyword. The cloud server then maps the keyword onto the set of matching data files and returns the file according to the keyword. The cloud system returns the data files on the basis of two rules: a) The user's input is precisely coordinates with the predefined keyword on the cloud and b) if there are some typographical errors or some format irregularities, then the server returns the closest matching results based on the predefined semantics to find the similarity. Following figure shows how the system works.

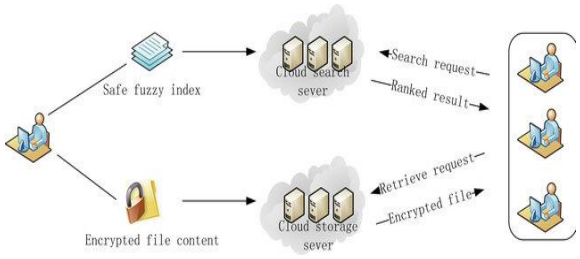

Fig 1. System Architecture

3.2 Design Goals

Fuzzy keyword search woks on accomplishing the two major requirements of constructing fuzzy keywords for searching a.) The methodology designed should be secure and data integrity must never be compromised. b.) Fuzzy keyword sets that are made must not only identify the exactly same keyword, but they must also identify keywords with slight inconsistencies like format errors or typographical errors.

So, to create storage and security effective fuzzy keyword search technique another technique is used which is 'wildcard based technique'. All the edit operations on the keywords for construction of fuzzy keyword sets are performed by using the edit distance operation of wildcard based technique.

For example, let there be two words w1 and w2. So, the edit distance between w1 and w2 would be represented as ed(w1,w2) and would be defined as the total number of operations needed to be performed on completely transforming one into the other. There are three operations related to the edit distance operation; a) Substitution b) Deletion and c) Insertion. The working of these three operations is quite clear from the names of the operations. Substitution is changing a letter in a word with another letter. Insertion is adding a character(s) in a word and Deletion is removing a character(s) from a word.

For example, for the keyword FEVER let the pre-set edit distance be 1, so, its fuzzy keyword set on the basis of wildcard technique can be made as

ed(FEVER,1) = {FEVER, *FEVER, *EVER, FE*VER, F*VER, FEV*ER, FEV*R, FEVE*R, FEVE*, FEVER*}

Reduction of more storage overhead can be done, with greater the pre-set edit distance: this wildcard based technique can also help in reducing the index storage to 40MB from 30GB approximately. In case the edit distance is set to be 2 or 3, the number is only O(l^d) for the keyword with length l and edit distance d.

4. SYSTEM CONSTRUCTION

Fig.1 illustrates the exact functioning of the technique suggested for performing effective fuzzy keyword search. Following is a detailed explanation of the construction of the once system

a) Registration

First of all, the user initiates a request to the owner of the cloud to register on the network. The owner then allows the user to register as a new user on the network. Once the user is registered, then he/she can easily login by providing the login credentials. All the users are managed by the owner and the credentials are matched from those saved in the database.

b) File Upload

A registered user can upload any file which follows the guidelines set by the owner. While uploading a file, a keyword is to be specified for the file which is used for searching the file. The user can upload as many files as he/she wants but each and every file must have a unique keyword.

c) Encryption

As soon as the file is uploaded, the system encrypts the keyword specified by the user using AES (Advanced Encryption Standards) algorithm. AES is based on Rijndeal Cipher and works on the substitution permutation network principle and uses a 128-bit key for encryption. All these features make AES extremely secure and accomplishes the first requirement of searching that is 'security must never be compromised'.

After the file is uploaded, the system automatically generates the fuzzy keyword sets as per the keyword used for storing the file. These sets are made by the help of wildcard based technique. [4]

d) File Search

A user can search a file by the help of the keyword used to upload the file. As soon as the user enters the keyword, the system again uses the edit distance operation to calculate the distance and match it with that of the files stored on the server. After this, the server sends the files that match with the keyword provided. System then decrypts the results that are provided by the system. The user, hence gets the relevant files and can easily choose the file he required from the provided list of files.

The above mentioned method provides an effective way of preserving data privacy and maintains non-impersonation efficiently.

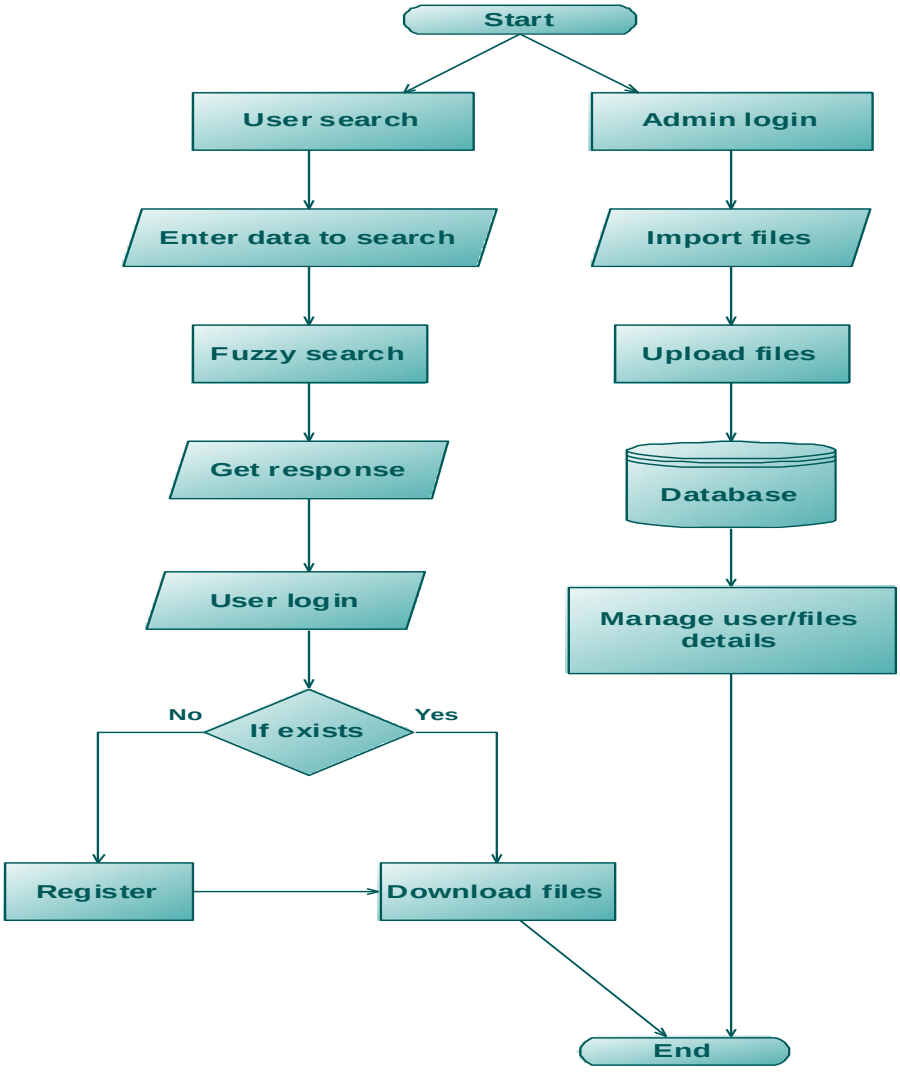

Fig 2: Data Flow Diagram of the Application

5. RELIABILITY ANALYSIS

The most important factor in web services is Reliability. The users must be able to access their data anytime, anywhere without any failure.

The suggested technique also fulfils this major requirement of web services. Factors affecting the reliability of data in the suggested system are:

5.1 **Availability**: As the system is online, the user can access it anytime, anywhere without any failure. User can access it either through his/her computers when at home or through handheld devices when on the go. The user must only have a stable internet connection to access the application.

5.2 **Performance**: The high-end performance of the application is ensured by the use of searchable encryption scheme and fuzzy keyword set construction. This has resulted in an extreme reduction in the time and space required to perform search operations and store the fuzzy keyword and has resulted in increasing the performance of the system.[3]

5.3 **Security**: As the data is encrypted prior to uploading it on the server, there is no issue regarding the security of the data. Also, the admin and user integrity is preserved by the help of admin and user authentication.

6. CONCLUSION

Introduction of cloud computing has completely changed the way of data storage in the modern era. With advancements of cloud, more and more users are preferring cloud over traditional storage techniques. With its help, users can store their files on the cloud and access them whenever and wherever they want. The proposed system helps in accomplishing the task in a simple and yet an effective manner. It performs a privacy-preserving and efficient fuzzy search over the data stored on a cloud. Other techniques like wildcard based technique is used to create fuzzy keyword sets effectively for searching the file. The sets are then used to perform a fuzzy keyword search over the encrypted data using the scheme proposed in this paper. The solution proves to be enhance security and yet preserve privacy. It also enhances the overall performance by increasing the speed of the operation, hence accomplishing the goal of fuzzy keyword search.

7. ACKNOWLEDGMENTS

We would like to thank Department of Computer Engineering, Amity School of Engineering and Technology and our guide Prof. Darothi Sarkar for her continuous help and support which helped us to design this paper.

8. REFERENCES

[1] Techniques for Efficient Keyword Search in Cloud Computing , P.Niranjan Reddy,Y.Swetha ,Department of CSE , Kakatiya Institute Of Technology & Science, Warangal Dist-506002,India.

[2] Fuzzy Keyword Search over Encrypted Data in Cloud Computing Jin Li, Qian Wang, Cong Wang†, Ning Cao ‡, Kui Ren †, and Wenjing Lou‡ †Department of ECE, Illinois Institute of Technology ‡Department of ECE, Worcester Polytechnic Institute Email: †{jinli, qian, cong, kren}@ece.iit.edu, ‡{ncao, wjlou}@ece.wpi.edu Secure Fuzzy keyword Search using an Advanced Technique over Encrypted Cloud Data

[3] Deeptha Hegde, Saritha Department of Computer Science and Engineering Sahyadri College of Engineering and Management Mangalore-575007 Efficient Implementation of AES

[4] Ritu Pahal Vikas kumar Dept. ECE, SGI Samalkha, Haryana, India SGI Samalkha, Haryana, India

[5] D.Boneh, G. Di Crescenzo, R. Ostrovsky, G. Persiano. ”Public key encryption with keyword search”, Proc. Of EUROCRYPT, 2004, pp.506-522.

[6] Sameer Rajan , Apurva Jairath.Cloud Computing. 2011. The Fifth generation of Computing, International Conference on Communication Systems and Network Technologies.